

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:04:33 ; Search time 18 Seconds

(without alignments)  
1276.451 Million cell updates/sec

Title: US-09-887-784-4

Perfect score: 1274

Sequence: 1 MVSNGELFTGVPIVLVD.....VLLGFVTAAGTILGNDELK 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR73.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1241	97.4	238	1 JQ1514	green-fluorescent
2	105	8.2	785	2 H72228	hypothetical prote
3	91.5	7.2	861	2 H61102	luciferin-ATP ligase
4	89.0	7.1	887	2 E82590	leucodextrin protei
5	89.5	6.9	655	2 D83917	DNA topoisomerase
6	87.5	6.9	370	2 E70390	iron-sulfur cofact
7	87.5	6.9	860	2 AC0582	leucyl-L-arginine synthet
8	87.5	6.9	2232	1 A36028	DNA-directed DNA p
9	87.5	6.9	2573	2 D71614	hypothetical prote
10	87.5	6.9	2573	2 D71614	hypothetical prote
11	87	6.8	874	1 J40930	glycolipase d
12	87	6.8	874	1 J40930	glycolipase d
13	86.5	6.8	797	2 JCA078	protective surface
14	86.5	6.8	808	2 F64102	RNA (uracil-5'-) m
15	86	6.8	357	2 G81355	DNA-directed DNA p
16	85.5	6.7	788	1 JCVLH8	inter-alpha-trypsin
17	85.5	6.7	899	2 G35576	cellulase (EC 3.2.
18	85.3	6.7	541	2 S25012	hypothetical prote
19	85	6.7	439	2 JH0414	synaptotagmin p95
20	85	6.7	439	2 JH0414	synaptotagmin p95
21	84.5	6.6	425	2 C97354	hypothetical prote
22	84.5	6.6	613	2 A99532	oligonucleotidase
23	84	6.6	353	2 E84941	imidazoleglycerol-
24	83.5	6.6	836	1 JVDVLD	DNA-directed DNA p
25	83.5	6.6	1134	2 A60234	Iga Fc receptor pr
26	83	6.5	1461	2 T02936	hypothetical prote
27	83	6.5	1461	2 T02936	hypothetical prote
28	83	6.5	471	2 T27836	hypothetical prote
29	83	6.5	774	2 T39539	alpha-amylase homo

nitrogenase (EC 1.  
neuraminidase, pro  
hypothetical prote  
hypothetical prote  
concomitant with  
proprotein convert  
lipoygenase (EC 1  
water-stress-induc  
synergohemotropi  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
coatomer complex a  
conserved hypothet

## ALIGNMENTS

### RESULT 1

JQ1514

green-fluorescent protein [valuated] - hydromedusa (Aequorea victoria)  
C:Species: Aequorea victoria  
C:Date: 03-Dec-1999, Revision: 03-Dec-1999, #text change: 23-Mar-2001  
C:Accession: P00692, GI:54693, EMBL:54693, F00335, S51330, S51331  
R:Praspher, D.C.; Skenorode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.  
Gene 111, 229-233, 1992  
A>Title: Primary structure of the Aequorea victoria green-fluorescent protein.  
A:Reference number: JQ1514; MUID:92175527; PMID:1347277

A:Accession: JQ1514

A:Molecule type: DNA

A:Cross-references: GB:M62654; NID:9155662; PID:AAA27722.1; PID:g155663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99, F, 101-140, L, 142-218, V, 220-238 <PRA>

A:Cross-references: GB:M62653; NID:9155660; PID:AAA27721.1; PID:g155661

A:Accession: P00335

A:Molecule type: Protein

A:Residues: 146-54, 74-122, 132-151, 154-183, 185-200 <PRA>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A>Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A:Molecule type: Preliminary

A:Residues: 1-24, O, 26-156, P, 158-171, K, 173-238 <INO>

A:Cross-references: GB:L29345; NID:9606383; PID:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13, V, 15-24, O, 26-44, N, 46-153, G, 155-156, P, 158-171, K, 173-227

A:Cross-references: EMBL:X83959; NID:9634008; PID:CAA58789.1; PID:g634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24, G, 26-29, B, 31-83, L, 85-153, G, 155-156, P, 158-171, K, 173-208

A:Cross-references: EMBL:G634010; NID:9634010; PID:CAA58790.1; PID:g634011

A:Experimental source: clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation: X-ray crystallography, 1.9 angstroms. residues 'A', 2-79, 'R'

A>Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia cc

Watkins, J.N.; Moss, L.G.; Phillips Jr., G.N.

A>Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98284543; PMID:9631087

[REDACTED]

---

Db 164 KGVPLLDVAQAIKPIELANISVATSGHKFHAIGSGELYISDEANVEPLVGGQE 223  
51 TGKLP- ----VDPWPLVTLLSYGVQCFSPYDHHKQ-HDFKFSAMPEGTVQETIFFKDD. 104  
QY

A.Reference: S60930  
A.Accession: S60919  
A.Molecule type: DNA  
A.Residues: 1-2221 <SEN>  
A.Cross-references: EMBL:X92494; NID:g10452326; PIDN:CAAG63235.1; PID:g1045247  
R:Son-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.  
submitted to the Protein Sequence Database, April 1996  
A.Reference: S63235  
A.Accession: S63235  
A.Molecule type: DNA  
A.Residues: 1-2222 <SEN>  
A.Cross-references: EMBL:Z71538; NID:g1302316; PIDN:CAAG96169.1; PID:g1302317; CSPDB  
A.Experimental source: strain S288C  
R:Son-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.  
Yeast 12, 505-514, 1996  
A.Reference: S65111  
A.Accession: S65111  
A.Molecule type: DNA  
A.Status: nucleic acid sequence not shown; translation not shown  
A.Cross-references: EMBL:X92494; NID:g10452326; PIDN:CAAG63235.1; PID:g1045247  
A.Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1999  
A.Gene: SGD:POL2; DUN2: MIPS:YNL262w  
A.Cross-references: SGD:S0005206; MIPS:YNL262w  
A.Map position: 14L  
A.Superfamily: DNA-directed DNA polymerase II  
C:Keyword: DNA binding; nucleotidyltransferase; nucleus; zinc finger  
Query Match  
Local Similarity 26 94; Score 87.5; D8 1; Length 2222;  
Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;

QY	54	LPVPMPLVLTVTYSVOCFSRIPDH-----KQDFKSAKMPGV---QERTI	99
		: : :     : : :     : : :	
DB	883	LPASPEVTFYFLANGKYLITSPCSMLNRVHOKTTHNQVQLMDPLANIYETHSENTI	942
		: : :     : : :     : : :	
QY	100	PFSDGQNYKTR--AEYKPEGDTILNR-----TELKIDFKEOGALGHKLKYNV	147
		: : :     : : :     : : :	
QY	943	FEVDGPPYKATILPSSKESGAKIKRVAVFNDGSLAGFLARRGEL---OLIKNFQ	999
		: : :     : : :     : : :	
QY	148	S--HNYIMAD	156
		: : :     : : :     : : :	
DB	1000	SDIRKFLVLEGD	1010
		: : :     : : :     : : :	
RESULT 10			
D71614			
hypothetical protein PF90460c - malaria parasite (Plasmodium falciparum)			

C>Date: 13-Nov-1998 revision 13-Nov-1998 text\_change 21-Jul-2000  
 C:Accession: D71614  
 R:Character: Nucleotide  
 K:Gardner, M.J.; Saito, S.; Zhou, L.; Sutton, G.G.; Clayton, K.; White, O.; Smith,  
 S. Science 287, 1126-1132, 1998  
 A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; MUID:99021743; PMID:9804551  
 A:Accession: D71614  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2573 <GAP>  
 A:Cross-references: GB:AF001362; GB:AF001366; GB:AF001388; PIDN:AC71881.1; PID:938  
 A:Experimental source: Clone 3D7  
 C:Genetic:

A7Gene: PF08460C

	Query Match	6.9%	Score 87.5;	DB 2:	Length 2573;
	Best Local Similarity	26.1%;	Pred. No. 1.2e+02;		
	Matches 34; Conservative	30; Mismatches	53; Indels	13; Gaps	5;
OY	94 QOERTIFPKD--DGNTKTAFAVFGDTLVNRIELAGIDFKGNILGHLEYN--VNSH 149			: :	:

A:Residues: 1-874 <ZHU>  
A:Cross-references: GB:U38842; NID:G1055336; PIDN:AA044405.1; PID:G1055337  
A:Experimental source: strain NM105  
C:Comment: This protein is a glycoprotein. It functions as protective coats, molecule  
C:Genetics:  
A:Gene: o1pA  
C:Superfamily: S-layer repeat homology  
F:30-974/Product: S-layer sequence #status predicted <SIG>  
F:157-209/Domain: S-layer repeat homology <SLR3>  
Query Match 6.8%; Score 87; D: 32; Length 874;  
Best Local Similarity 22.8%; Pred. No. 32;  
Matches 65; Conservative 28; Mismatches 96; Indels 96; Gaps 14;  
QY 6 EELP--TGVPPII---VELGDVNGKHSVSGE3GDA-----YKGLTKFI 48  
DB 551 KVFPGKIVGLVLDVTTTNGSIGCTSIKYKNGVAGTIFQNPASGEGYSLRVEVT 70  
QY 49 CTGKPLVPVPTLV-----TTLSVGQCFSRYPDHMKOHDFKSNAPGEGYQVR 97  
DB 611 KNSITGEAPLELVSFKAGOGKGAEDATLTGAGNTVAYGLSNTTGGYVADADLAGY--EF 668  
QY 98 TFFKDDGNKYTRAEVDFEGDTLVNRIELKGD---FKFDGNLGHLEYNNSHYVIM 154  
DB 669 RV-----GNQKI--ASAKIGKTLKVGTYAGVYDILDKGTAGTATVYQENIQIT 721  
QY 155 AKDKK-----NGIKVNSKIEHIN-----IEDGSVOLADH 182  
DB 772 SVKPKVEVEQFENKYNIDRVLDVYKSDKDDVLNGIKLNISTEHRKIVDEGTQ----- 777  
QY 183 YQONTPIGDGVPVLDPNHYLSITQSALSXDPNEKRQDHVLLGSPYA 227  
DB 778 -----GKV-----YLDKDNQATFDGND-----VALGVTYA 802  
RESULT 13  
QY 78 JCA078  
C:Species: Haemophilus influenzae  
A:Variety: type b  
C:Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
C:Accession: JCA078  
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.  
Gene 156, 97-99, 1995  
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae JCA078  
A:Reference number: JCA078; NID:95255676; PMID:7737523  
A:Molecule: type: DNA  
A:Residues: 1-797 <FLA>  
A:Cross-references: GB:U13961; NID:G537447; PIDN:AA85645.1; PID:G537448  
A:Experimental source: type b  
C:Superfamily: protective surface antigen D-15  
C:Keywords: surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>  
Query Match 6.8%; Score 86.5; DB 2; Length 797;  
Best Local Similarity 21.9%; Pred. No. 32;  
Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;  
QY 65 LSYGVQCFSRYPDHMKQHD-----FKSNAPGEGYQVR-----RTI 99  
DB 427 IGKTEGEGISTQASVKDNDFLGTGAANVSTATKNDYSGVYTPYTKDGVSGVGNV 486  
QY 100 FKDDGNKYTRAEVDFEGDTLVNRIELKIDFEGDNI---LGH-----KLEYNYS 148  
DB 487 FFENYDNKSQTSNNKYKTTGNSNTL-GFPVNNNSYVYGLHTYKISNFALEYN--- 542  
QY 149 HNYITADKQR-NGIKVNFRIHNEIDGVSOLADHYQO-----NPTIDGQVLL 196  
DB 543 RNLYTQSMFKNGCINKTN-----DFDFSEFGNNTNSLRGYVFPFKGVKASLG-GRVTI 593

